EFFECTS OF GAIN/LOSS FRAMING IN CYBER DEFENSE DECISION-MAKING

Nathan Bos[1], Celeste Lyn Paul[2], John R. Gersh[1], Ariel Greenberg[1], Christine Piatko[1], Scott Sperling[1], Jason Spitaletta[1], Dustin L. Arendt[3], Russ Burtner[3]

[1]Johns Hopkins University Applied Physics Laboratory
[2]U.S. Department of Defense
[3]Pacific Northwest National Laboratory

Cyber defense requires decision making under uncertainty, yet this critical area has not been a focus of research in judgment and decision-making. Future defense systems, which will rely on software-defined networks and may employ "moving target" defenses, will increasingly automate lower level detection and analysis, but will still require humans in the loop for higher level judgment. We studied the decision making process and outcomes of 17 experienced network defense professionals who worked through a set of realistic network defense scenarios. We manipulated gain versus loss framing in a cyber defense scenario, and found significant effects in one of two focal problems. Defenders that began with a network already in quarantine (gain framing) used a quarantine system more, as measured by cost, than those that did not (loss framing). We also found some difference in perceived workload and efficacy. Alternate explanations of these findings and implications for network defense are discussed.

INTRODUCTION

Cyber defense requires decision making under uncertainty (Tversky & Kahneman, 1974). Defenders must synthesize and interpret evidence from multiple imperfect sensors, conduct investigations that entail time and resource costs, and make decisions about deploying defensive measures that may also have time and resource costs for both the analytic team and the organization being defended. Cyber defenders make decisions based on risk/reward calculations and uncertain information. The paucity of research in this area stands in contrast with the strong body of research on sensemaking in medicine, emergency response, financial analysis, and similar fields where real-time responses based on uncertain information are required (e.g., Klein, 1998).

Modern Cyber Defense

The focus of cyber defense work involves more judgment than it has in the past as capabilities increase and the functions of detection and analysis become increasingly automated. D'Amico and Kocka's 2005 workflow analysis of network defenders identified three phases of the workflow: detection, situation assessment and threat assessment. This early research remains one of the most thorough workflow studies. However, a more recent analysis includes additional stages reflecting higher level judgments. Reed, et al's (2014) first three stages match D'Amico and Kocka's fairly well: Review alerts (detection), Understand (situation assessment), and Evaluate risk (threat assessment). Their model adds Mitigation and Deeper Dive. Mitigation includes actions such as blocking, killing processes, and policy changes, which are included in the quarantine actions of the current study. The need for a Deeper Dive stage may explain some of the shortcomings of analyses we observed. The first stages, detection and analysis, are increasingly automated. Network defenders we have interviewed describe how a decade ago they spent more time analyzing network traffic logs, whereby today's defenders spend more time investigating automated alerts (Gersh and Bos, 2014).

New tools can offer much greater power in the evaluation and mitigation stages. Software-defined networking, for example, will soon allow a defender at a single workstation to perform mitigation actions within seconds for operations that might have required hours of onsite, hands-on work with workstations and network hardware.

Moving Target Defense (MTD) is another development that would require rapid high level judgments from defenders. MTD's change the properties and structure of a network in real-time to make it a more difficult target to attack. MTD is often described in the context of five domains: Data, Software, Runtime Environment, Platform, and Networks (e.g., Carvalho et al., 2012; Hobson et al., 2014). Techniques in this domain often require human-in-the-loop in order to operate and be effective. However, the relative newness of the MTD research area means there is a lack of understanding how these techniques may affect cyber defenders' comprehension and performance.

New automated tools often put humans in a position of supervising automation and applying human judgment. A typical situation may be when an automated defense system perceives a threat and enacts a mitigation such as throttling (slowing) an upload in progress, blocking an external website, or even isolating a user computer. The defenders would then apply human judgment as to whether to keep, extend, or undo the automated actions. This decision then becomes one of weighing potential gains versus losses in the face of uncertainty.

Judgment and Decision Making

While the technologies are new, the difficult judgment problems they present are familiar. Judgment and decision-making, which focuses on decision made from uncertain

information, involving complex risks, and time-constrained decision making, is an active subfield of cognitive psychology, much of it built around the framework laid by Tversky and Kahneman (1974; 1979). Their early findings on risk judgments are collectively referred to a *prospect theory* (1979). Gain versus loss framing is a central component of prospect theory. Tversky and Kahneman (1991) unpack this phenomenon into three components: loss aversion, reference points, and diminished sensitivity.

Loss aversion. Humans tend to avoid perceived losses more than they seek perceived gains. This is often demonstrated in simple studies by giving participants a small gift such as a mug, and later comparing the dollar value owners place on the mugs to the price other participants would be willing to forego for the same object. Ownership seems to almost instantly increase perceived value of assets; in the 1991 study, owners valued newly-acquired mugs at around $7, while potential buyers valued them at just over $3. This is sometimes considered a special case of a *status quo bias*.

Reference points. Humans judge gains and losses from the perspective of their current state, or perceived current state. In the mug example, ownership of the mug is a reference point. For many examples, financial and non-financial, moving an imagined reference state can change what is framed as a loss versus gain and thus influence decisions.

Diminished sensitivity. Humans are most sensitive to small changes when they differ from the status quo. Large differences, however, are not valued proportionately higher than small differences. The difference between a $0 and $1 loss is much more salient than the difference between a $10 and $11 loss.

Uncertainty also strongly influences decision making, in ways that interact with gain/loss framing. When faced with potential gains, humans often choose a smaller, certain gain over a larger uncertain gain, e.g. a person may choose a $100 certain payoff versus a 50% change to win $210, which has an expected value of $105. This can be reversed for losses, where people would tend to choose a 50% risk of losing $210 rather than accept a certain loss of $100. However, the perceived likelihood of these gains and losses also interact with gain/loss framing, whereby very high and very low probability changes can be treated differently. Uncertainty was not manipulated in our study, but may help interpret and understand our results.

Previous decision-making research in Cyber Defense

User decision making. There is a body of research on how users make decisions affecting security. One example is Rossoff, Cui and John (2013), who studied user decisions to make risky choices online. The authors found that risky decisions declined when users were prompted to recall friends' prior negative experiences online. The authors also implemented a simple gain vs. loss framing by rephrasing of a line in the instructions. A gain phrasing was "If she presses 'do not proceed,' she may avoid the risk of acquiring a virus that will cause serious damage to her computer." The corresponding loss framing was "If she presses 'proceed,' she may risk acquiring a virus that will cause serious damage to her computer." This gain/loss framing did have a small effect on behavior; participants were more likely to make a risky decision in the gain condition.

Institutional projection. Another line of research that makes use of what is known about risk and decision making looks at institutional investment in cyber protection. Öğüt et al. (2011) studied institutional investment in cyber defense. They concluded that most institutions did not invest in defense commensurate with the risk and cost incurred. Generally institutions only invested what the authors judged to be an appropriate amount when required to do so by insurance companies. This finding is consistent with research on human decision-making, where humans often choose lower-utility uncertain outcomes rather than pay sure costs.

METHODOLOGY

Our work aims to understand how judgments of risk and resulting decisions of cyber defenders are affected in a dynamic network environment, thus leading to our hypothesis:

The initial status of a network (completely isolated or fully connected) has an effect on quarantine decisions made by cyber defenders.

Participants were asked to respond to a cyber defense scenario and quarantine affected computers with the goal of minimizing risks to the network while balancing business costs. The starting network condition (isolated or connected) was a counterbalanced between subject design.

The design of the experiment was informed by previous related research activities including observations of professional network defenders (Gersh and Bos, 2014), a pilot study with cyber security graduate students (APL, 2014), and expert interviews with professional cyber security experts (APL, 2014).

This study was reviewed and approved by PNNL's Institutional Review Board.

Participants

17 participants with professional cyber security experience were recruited through word of mouth and direct contact from a research laboratory. One additional participant was recruited but did not have the requisite experience to complete the study. Participants had a range of skills and experiences including computer security incident responder (n=11), cyber threat analyst (n=9), security administrator (n=9), security consultant (n=9), and other defensive cyber roles (n=10). The median participant experience in cyber security was 5 to 10 years (n=8).

Network Quarantine System

Ocelot (Arendt et al. 2015) is a visual interface that supports dynamic network management. Specifically it supports visualization, analysis, and mitigation.

Visualization uses a novel hierarchical approach of flexibly grouping network devices according to arbitrary attributes that represent connections as links between nodes. A timeline visualization is integrated with the main visualization to show network traffic over time. The user can use the traffic

visualization to learn when individual connections between devices were active.

Analysis is supported through query and filtering tools and set-based operations. Rather than being limited to a static, topology based diagram, the defender can flexibly group and filter based on any attribute, and show/hide connections between arbitrary sets of nodes.

Mitigation is accomplished through a quarantine system. (The effects are visualized in the interface even though there is no network.) Quarantine actions include external isolation (e.g., block connections to Internet), internal isolation (e.g., block connections to intranet computers and services), throttle network traffic (e.g., slow outbound traffic), recredential users (e.g., force password reset), push software patches (e.g., Flash), additional monitoring (e.g., logging). More than one quarantine option could be applied to a network set.

Cyber Defense Scenario

The experimental scenario was based on the IEEE Visual Analyics Science and Technology (VAST) 2013 mini-challenge 3 dataset (Whiting et al., 2013) that we modified to meet our experimental requirements. Participants took on the role of a network defender called in to assist the 'Big Marketing' advertisement company. The corporate network consisted of 1100 computers, including user workstations that employees used for day-to-day work, plus servers that supported critical business services and store sensitive business information. Three plausible threats were described for the participants, including a hacktivist organization intent on embarrassing a key client, a disgruntled former employee, and corporate competitors.

Over the course of one "week" (simulated in a 2 hour study session), participants were asked to respond to a series of cyber incidents that occurred on the network. Participants were asked to analyze the attack and decide at what level to quarantine a subset of computers. They then had to decide how to respond to the attack by applying quarantine options to the affected computers, balancing the cost of taking business services down with the costs of a continuous advanced persistent threat.

Three cyber events embedded in the VAST 2013 scenario were isolated and presented as problems. The first, a denial of service attack on a corporate server, was used for training. The following problems were used as research data.

Problem 1 (Redirecting web server). A corporate server has been intermittently redirecting traffic to an external site. Defenders could perform analysis by using set operations to identify vulnerable machines and link analysis to determine which users had been redirected. Defenders then could use the mitigation functions along with their judgment in deciding what to do with the compromised server, deciding whether to take proactive action with a server displaying similar vulnerability, and deciding what to do with redirected internal computers that may have been compromised through a Flash vulnerability.

Problem 2 (Zombie botnet). Big Marketing user machines have been part of a DoS attack launched against a customer website. Defenders could use the timeline to identify the time of attack, the timeline plus set operations to separate attackers from normal traffic, and set operations to identify a likely common vulnerability (unpatched Java.) Defender then could use the mitigation system with their own judgment to decide how to deal with Botnet victims, the external site, and other user machines with similar vulnerabilities.

Isolated vs. Connected Gain/Loss Manipulation

In half of the problems, the starting state had the network in complete isolation (net-down), as explained with this text:

A few hours before you arrive the CEO, alarmed at this news, ordered all network services be shut down pending your arrival, including external, internal, email, printing and fileserver access. He has given you full authority to decide what level of quarantines to remove and what if any to keep. This manipulation created the gain/loss framing. With the entire network already in quarantine, every machine that was changed to a less restrictive state represented a sure gain. In the opposite starting state, every machine that was quarantined represented a sure loss. The potential risks of compromise and exfiltration were uncertain losses. The conditions were counterbalanced such that half of the participants started Problem 1 with the network isolated and Problem 2 with the network connected; the other half had the opposite starting states. Because the Training problem and Problems 1 and 2 was a single narrative where the second problem built upon the first, the order that the problems were presented to participants could not be randomized.

Self Assessments

Several self assessments were administered after each problem to measure risk, decision efficacy and cognitive load. The rating scale of these assessments was normalized to a 7 point scale with high, low, and neutral anchor points.

Risk Assessment. Participants were asked to assess the likelihood of several risks based on the problem they completed (summarized): *Attack is conducted by sophisticated attacker, Incident is a false alarm, Incident will be handled quickly, Incident will lead to worse if not addressed, Incident is part of larger threat*.

Decision Efficacy. Participants were asked to rate their agreement on the effectiveness of their decisions (summarized): *I made the right decision; My decisions prevented harm to the network; My decisions kept the network safely active.*

Perceived Workload. A subset of items from the NASA TLX were adapted to create a measure of perceived workload. Time pressure and physical demand were omitted from the scale, and all items were put on a 1-7 scale. An additional measure for problem difficulty was added.

Cost of Quarantine Actions

Throughout the scenario, participants were instructed to consider the following business impacts to support their risk analysis during network quarantine actions (Table 1).

Table 1: Business impact of lost network services per day

Network Service	Cost/day
Corporate website (web)	$10,000
Sales portal (web)	$22,000
Customer service portal (web)	$80,000
Fileserver 1	$12,000
Fileserver 2	$20,000
Fileserver 3	$20,000
Access to external Internet, per website. Cost per user per day = $41.	$15,000
Access to intranet, including fileserver, email etc. Cost per user per day = $95	$35,000
Total cost of network service, Internet and intranet	$262,000

The impacts of different quarantine actions applied to network services were also provided (Table 2). These impacts were developed through expert interviews with subject matter experts.

Table 2: Business impact of quarantine actions

Quarantine Actions	Cost Impact
Internal and external isolation	100.0%
Internal isolation	75.0%
External isolation	65.0%
Throttle network traffic	37.5%
Recredential users	25.0%
Push multiple software patches	25.0%
Push one software patch	12.5%
Additional monitoring	12.5%

The cost of network services and cost impact of quarantine information were used to calculate the overall cost of quarantine decisions made by participants. For example, pushing a software patch to the corporate website (12.5% * $10,000) would result in a daily cost (loss) of $1250.

Procedure

The study was conducted in the following way. First informed consent was received from the participant. Then, background information on the network, estimated costs of network services, and a dossier of potential cyber threats to the company were provided to participants. Participants were then given an overview of the network quarantine system that included demonstrations of capabilities by the experimenter. Participants then used the system to complete a training problem guided by the experimenter.

Following the training, participants moved on to the main scenario problems. The experimenter did not provide analysis assistance but did provide usability assistance if the participant asked how to do something with the system. After each problem, participants completed a threat assessment, decision efficacy, and perceived workload assessment. At the end of each scenario problem, participants completed a general cyber risk assessment and system usability scale. Sessions lasted approximately 2 hours. In addition to the moderator, one to two observers took notes during each study session.

Analyses of within-subject variables were done with a paired t-test, and between subjects comparisons were done with a t test assuming equal variance.

RESULTS

We found significant effects ($p < .05$) for network gain/loss framing in both Problem 1 and Problem 2; however, the effects differed across problems.

Problem 1: Redirecting Web Server

There was a significant network framing effect for quarantine cost in Problem 1, $t(15) = 4.98$, $p < .001$, with the Isolated condition resulting in higher overall costs than the Connected condition (Figure 1). This fit the hypothesis that potential losses (new quarantines) would be avoided to a greater extent than gains (lifting quarantines) would be sought.

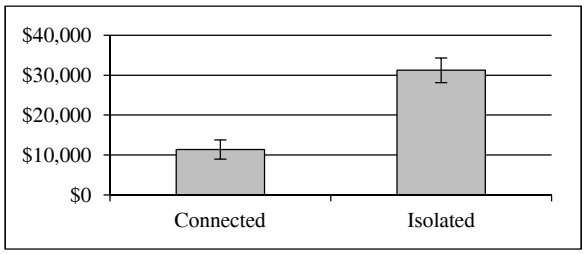

Figure 1: Cost of network quarantine actions across network framing conditions in Problem 1.

There were no significant framing effects in Problem 1 for the risk assessment, decision efficacy, perceive workload, or participant experience.

Problem 2: Zombie Botnet

There was no overall difference in the level of quarantine based on the gain/loss manipulation in problem 2. There were some differences in self-report responses. There was a significant network framing effect for decision efficacy in Problem 2, $t(15) = -2.20$, $p = .044$, with participants in the Isolated condition having higher confidence in their decisions to keep the network active (Figure 2). There were no effects for making the right decision or preventing harm to network.

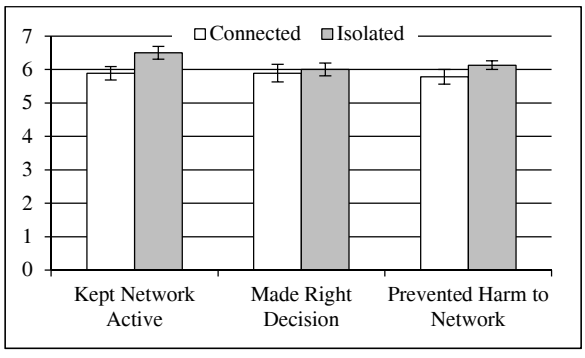

Figure 2: Differences in network framing conditions across Decision Efficacy questions in Problem 2.

There were also significant network framing effects for several workload measurements. The Isolated condition had higher perceived workload, resulting in higher Mental Demand ($t(15) = -1.97$, $p = .024$), Frustration ($t(15) = -1.81$, p

= .033), Work Hard ($t(15) = -1.81$, $p = .019$), and Stress ($t(15) = -1.92$, $p = .035$), but not for Task Difficulty (Figure 3).

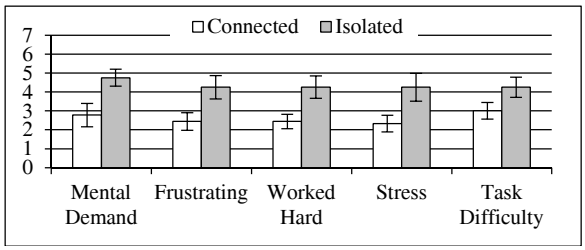

Figure 3: Differences in network framing conditions across workload questions in Problem 2.

There were no significant network framing effects in Problem 2 for the risk assessment.

DISCUSSION

Network Gain/Loss Framing

The experimental manipulation showed the expected pattern on one cyber defense problem (Redirect) but not the other (Botnet). Participants who started the Redirect problem with the network in Isolation made more costly quarantine decisions than those that did not. Does this mixed result mean the hypothesis was supported, or not? We tend to interpret this as support for several reasons.

Because of the counterbalanced within-subject design, and because the order of problems was not counterbalanced, all of the Connected participants in the second problem had experienced the Isolated manipulation in the first problem. So any lingering effects would tend to mask results in the second problem.

Task demands and training effects may have led participants to stop at a partial solution for the Botnet problem, which was identifying the Botnet without addressing the underlying vulnerabilities. The majority of participants enacted a quarantine only on the Botnet machines and the external target machine and did not address other machines with unpatched Java, or other similar vulnerabilities. Referring to the Reed, et al. (2014) framework, the participants stopped after mitigation and did not do the "deeper dive".

Why was this different between problems? First, in Problem 1 there was explicit prompting to consider other vulnerable classes. Second, time constraints of the experiment were a factor; by the conclusion of Problem 2 most participants were close to the end of the two hour session. For these reasons, we hypothesized that problem demands and the way the problem set was trained on may have reduced variance in the solution set, masking other effects.

Assuming the gain/loss difference observed in Problem 1 is a real effect, what are the implications for network defense? The first, most obvious implication is to pay attention to framing effects. This is relevant for training and for development of defense policies.

Automated network shutdowns may be an increasingly common feature of defense systems. Systems based on software-defined networking in particular may be given the capacity to automatically block external sites and IP ranges, or throttle activity based on suspicious activity patterns, e.g. a large upload after hours. These automated systems eventually have humans in the loop to apply judgment, perhaps to lift quarantines or to change policies after investigation. Systematic bias in defender judgment based on network state could be taken into account in design of such systems.

The differences in self-reported workload after Problem 2 are harder to interpret. We knew that we had presented defenders with a difficult analytic and judgment task, but had no expectation that the gain/loss framing would interact with workload in this way. Further study of these cognitive processes is warranted.

More generally, this research opens the door to a better understanding of how network defenders make decisions under uncertainty. Cyber defense is a unique and important type of professional role where analysts work in a very fast-changing environment, and this role deserves the kind of research attention paid to other important professionals. The judgment and decision making literature provides an entryway into understanding risk/reward judgments in particular.

REFERENCES

Arendt, D. L., R. Burtner, D. M. Best, N. D. Bos, J. R. Gersh, C. D. Piatko, and C. L. Paul. (2015). Ocelot: User-Centered Design of a Decision Support Visualization for Network Quarantine. *Proc. IEEE Symposium on Visualization for Cyber Security*, 1-8.

D'Amico, A. and M. Kocka. (2005). Information assurance visualizations for specific stages of situational awareness and intended uses: lessons learned. *Proc. IEEE Workshop on Visualization for Cyber Security*, 107-112.

Carvalho, M., Bradshaw, J.M., Bunch, L., Eskridge, T., Feltovich, P.J., Hoffman, R.R., and Kidwell, D. (2012). Command and Control Requirements for Moving-Target Defense. *Human-Centered Computing*, 27(3), 79-85.

Gersh, J. R. and Bos, N. (2014). Cognitive and Organizational Challenges of Big Data in Cyber Defense. *Proc. Workshop on Human Centered Big Data Research*, 4-8.

Hobson, T., Okhravi, H., Bigelow, D. Rudd, R., and Streilein, W. (2014). On the Challenges of Effective Movement. *Proc. ACM Workshop on Moving Target Defense*, 41-50.

Johns Hopkins University Applied Physics Laboratory. (2014). Observations of Watchfloor Cyber Defenders, JHU/APL AOS-14-1089.

Klein, G. A. (1998). *Sources of Power: How people make decisions*. Cambridge, MA: MIT Press.

Öğüt,H., Raghunathan, S., and Menon , N. (2011). Cyber Security Risk Management: Public Policy Implications of Correlated Risk, Imperfect Ability to Prove Loss, and Observability of Self-Protection. *Risk Analysis*, 31(3), 497-512.

Reed, T., Abbott, R.G., Anderson, B., Nauer, K., and Forsythe, C. (2014). Simulation of Workflow and Threat Characteristics for Cyber Security Incident Response Teams. *Proc. Human Factors and Ergonomics Society Annual Meeting*, 427-431.

Rosoff, H., Cui, J., and John, R.S. (2013). Heuristics and biases in cyber security dilemmas. *Environmental Systems Decision,* 33, 517-529

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.

Tversky, A. and Kahneman, D. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 47(2), 263-291.

Tversky, A. and Kahneman, D., (1991). Loss aversion in riskless choice: a reference-dependent model. *The Quarterly Journal of Economics*, 106(4), 1039-1061.

Whiting, M., Cook, K., Paul, C.L., Whitley, K., Grinstein, G., Nebesh, B., Liggett, K., Cooper, M., and Fallon, J. (2013). VAST Challenge 2013: Situation Awareness and Prospective Analysis. *Proc. IEEE Conference on Visual Analytics Science and Technology.*